



Audio Cues to Assist Visual Search in Robotic System Operator Control Unit Displays

**by Ellen C. Haas, Ramakrishna S. Pillalamarri,
Christopher C. Stachowiak, and Michael A. Lattin**

ARL-TR-3632

December 2005

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

DESTRUCTION NOTICE—Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5425

ARL-TR-3632**December 2005**

Audio Cues to Assist Visual Search in Robotic System Operator Control Unit Displays

**Ellen C. Haas, Ramakrishna S. Pillalamarri, Christopher C. Stachowiak,
and Michael A. Lattin**
Human Research and Engineering Directorate, ARL

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<p>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) December 2005		2. REPORT TYPE Final		3. DATES COVERED (From - To) FY 2005	
4. TITLE AND SUBTITLE Audio Cues to Assist Visual Search in Robotic System Operator Control Unit Displays				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Ellen C. Haas, Ramakrishna S. Pillalamarri, Christopher C. Stachowiak, and Michael A. Lattin (all of ARL)				5d. PROJECT NUMBER 62716AH70	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory Human Research and Engineering Directorate Aberdeen Proving Ground, MD 21005-5425				8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-3632	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT <p>Mission demands have made the robotics collaboration operator control unit (OCU) into a relatively dynamic, demanding, cognitively complex system where Soldiers must perform multiple tasks such as controlling multiple robots and processing large amounts of information in environments that sometimes contain high levels of noise. Research and modeling data indicate that audio display technologies would be very useful in OCU applications such as guiding visual display search. The purpose of this study was to examine the effectiveness of the integration of auditory display technologies in visual search tasks such as those that occur in robotic OCUs. Independent variables were audio signal mapping scheme, type of verbal positional cue, and visual target azimuth. Dependent variables were visual target search time and the National Aeronautics and Space Administration Task Load Index workload rating of the target search task. Participants were 36 students (15 males and 21 females) from Harford Community College. The results indicated that the use of auditory signal mapping and verbal positional cues significantly reduced visual display search time and workload and that positional cues mixed with specific audio mappings were the most efficient means of reducing search time. Specific design recommendations are made regarding the use of auditory signals in environments with narrow field-of-view visual displays.</p>					
15. SUBJECT TERMS auditory assisted visual search; auditory displays					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 25	19a. NAME OF RESPONSIBLE PERSON Ellen C. Haas
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) 410-278-5825

Contents

List of Figures	iv
List of Tables	iv
1. Background	1
1.1 Audio Mapping	2
1.2 Verbal Positional Cues	3
2. Method	4
2.1 Participants	4
2.2 Apparatus.....	5
2.3 Stimuli	6
2.3.1 Visual Stimuli.....	6
2.3.2 Auditory Stimuli.....	6
2.4 Procedure.....	7
2.5 Experimental Design	8
2.6 Data Analysis	9
3. Results	10
3.1 Visual Target Search Time.....	10
4. Discussion	13
4.1 Target Search Times.....	13
4.2 Workload.....	15
5. Conclusions	16
6. References	17
Appendix A. Williams Square Designs Used in This Study	19
Distribution List	20

List of Figures

Figure 1. Participant with visual and auditory task apparatus.	5
Figure 2. Mean target search times for audio mapping x positional information, visual signals at 0 degrees absolute azimuth (+ 1 standard error of the mean).	11
Figure 3. Mean target search times for audio mapping x positional information, visual signals at 7.74 degrees absolute azimuth (+ 1 standard error of the mean).	11
Figure 4. Mean target search times for audio mapping x positional information, visual signals at 15 degrees absolute azimuth (+ 1 standard error of the mean).	12
Figure 5. Mean workload ratings for Audio Mapping x Positional Information interaction (+ 1 standard error of the mean).	13

List of Tables

Table A-1. Williams square design, order of presentation of auditory mapping scheme to participants.	19
Table A-2. Williams square design, order of presentation of visual target stimuli to participants.	19

1. Background

The U.S. Army is exploring the insertion of multiple advanced technologies, such as spatial audio displays, into a robotics collaboration operator control unit (OCU). Mission demands have made the OCU into a relatively dynamic, cognitively complex system where Soldiers must perform multiple tasks such as controlling multiple robots and processing large amounts of information in environments that sometimes contain high levels of noise. These demanding, cognitively complex tasks may result in Soldier information overload as well as decrements in situational awareness and mission performance. In addition, a human factors analysis of current robotic OCU systems (Bodenhamer, 2004) reveals that OCU visual display design and accompanying system warnings are less than adequate, with display design hindering search and poor use of cues. In the MATILDA¹ small unmanned robotic ground vehicle, OCU display quality was found to be poor, and the displays lacked system status or robot orientation indicators for the operator. The Buster unmanned aerial vehicle OCU lacked sufficient visual contrast to allow indication and interpretation of needed information and lacked appropriate warning indicators. Both MATILDA and Buster would benefit from auditory displays that supplement visual displays by providing information regarding system status, orientation, speed, or system warnings.

In the proposed robotics collaboration Army technology objective (ATO) systems and the Future Combat System (FCS), Soldiers will search visual displays for friendly and enemy objects and targets. Although there are no robotics collaboration platforms defined as of yet, it is possible that variations might use a standard 18-inch diagonal screen that produces a field of view (FOV) of approximately 30 degrees when viewed from a distance of approximately 2 feet (within the range of recommended viewing distances recommended in MIL-STD-1472F; U.S. Department of Defense, 1999). However, searching for and finding visual information within a 30-degree FOV might prove to be difficult, especially when there is a large number of visual objects on the display. Perrott, Sadralodabai, Saberi, and Strybel (1991) found that response time for target identification without audio cues doubled from 700 ms to 1400 ms when the number of visual objects increased from 2 to 64. Visual search may be very difficult in situations when visual targets include revisions of an already complex display, such as the change of military symbols on a map already populated by a large number of military symbols.

Research has shown that spatial audio cues can enhance visual search. Providing spatial auditory cues as guides (also known as aurally guided visual search) can facilitate detection of a visual target (Perrott et al. 1991; Elias, 1995, 1996; Perrott, Cisneros, McKinley, & D'Angelo, 1995,

¹Mesa Associates' Tactical Integrated Light Force Deployment Assembly

1996; Strybel, Boucher, Fujawa, & Volp, 1995; Fujawa & Strybel, 1997). Begault (1993) found that visual search times of commercial pilots were reduced by several (2.2) seconds when spatially correlated sounds were present. Elias (1996) found that aural cues enhanced visual search, particularly when 12 visual distractors were present. In two experiments, the slope of the function relating search time to number of distractors in an auditory spatial cueing condition was half that obtained with no auditory cue when targets are the same color as other items on the screen (Perrott et al., 1991; Strybel & Perrott, 1993). These data indicate that spatial audio displays may have considerable utility in guiding search of map displays.

Although audio cues have been found to be effective in guiding visual search, there has been little research to determine effective design characteristics of these cues. Two approaches to providing effective audio cues for visual search are mapping spatial audio signals to visual displays and adding verbal azimuth information to audio cues.

1.1 Audio Mapping

Mapping spatial audio information to visual displays with a 30-degree FOV is a challenging problem that few researchers have explored. Begault (1993, 1998) used a mapping scheme for a display with a 70-degree visual FOV in which spatial audio cues were presented in seven positions from 9 to 3 o'clock (a total of 180 degrees), with these positions exaggerated in relationship to the visual stimuli by a factor of two (i.e., visual targets at 15 degrees' azimuth would be signaled by an audio warning heard at 30 degrees' azimuth). Note that this exaggerated relationship would result in a more than 2-1 visual-to-audio-cue ratio when a visual cue at a 70-degree FOV would be heard at 180 degrees. Begault (1993) used discrete spatial audio positions rather than a continuous spatial mapping because he believed that discrete positions gave a more convincing suggestion of what direction to search. He also suggested that discrete spatial audio signal positions that exaggerate visual positions may be advantageous for several reasons: (a) exaggerated audio positions preclude the entire visual space from being mapped to only a limited spatial audio arc or angle; (b) discrimination error (as opposed to localization error) would be reduced with the use of exaggerated positions; and (c) the use of a greater range of azimuths would employ the full range of interaural time delays that the auditory system interprets in terms of left-right displacement. Although Begault (1998) ran a second study using another mapping scheme with a more limited audio range (audio alarms at eight positions from ± 60 degrees), he did not formally compare these two mapping schemes to each other, nor did he provide a suggestion of which was the more efficient.

A search of the literature indicates that neither Begault (1993, 1998) nor any other researcher suggested mapping schemes for displays with a FOV of less than 70 degrees. The purpose of this study was to determine whether efficient mapping schemes for displays with a narrower visual azimuth, such as displays with a 30-degree FOV (such as those found on potential robotic OCUs), could be developed. Five different strategies, three of which are based on Begault's recommendations regarding the use of discrete spatial audio positions and exaggerated audio

range, were compared in this study. The first strategy employed an audio display with three spatial audio positions at -15 , 0 , and $+15$ degrees' azimuth (a range of audio angles corresponding to the azimuth of the 30-degree FOV visual display). A second mapping strategy used an audio display with three spatial audio positions at -60 , 0 , and $+60$ degrees azimuth (a relatively wide range of audio angles). A third strategy used three spatial audio positions at -37 , 0 , and $+37$ degrees azimuth (approximately midway between ± 15 and ± 60 degrees). A fourth strategy employed a monaural presentation of audio cues, in which all cues would reach both ears at once, resulting in a nonspatial presentation in which all audio signals are heard as if coming from the center of the head (one audio position with no range of audio angles). Monaural audio was to be used because it is currently the most common mapping scheme used in audio displays. A fifth strategy was to use no sound at all (silence) as a baseline comparison condition and as a basis of comparison with displays that do not employ audio signals, such as displays with visual information only. It was hypothesized that because the first three mapping schemes use spatial audio, they would produce shorter visual search times than the monaural or the visual only schemes.

Comparing different spatial mapping schemes with monaural and visual only presentation will help define an effective means for mapping audio cues to visual targets within a relatively narrow range, such as a 30-degree FOV display. There is a possibility that the spatial mapping scheme that uses the widest range of audio angles (± 60 degrees) may produce significantly shorter response times than monaural displays and conditions with visual only but longer response times than displays with a narrower range of audio angles (± 37 or ± 15 degrees) because of a greater disparity between audio and visual cue. For example, an audio cue for a 15-degree visual target will be heard at 60 degrees for the ± 60 -degree audio display and at 37 degrees for the ± 37 -degree audio display. It is possible that for displays with a wider audio range, the larger disparity between audio and visual cue may confuse the listener, possibly increasing participant search time. However, the audio-visual cue disparity in Begault (1993, 1998) did not seem to interfere with participant performance.

1.2 Verbal Positional Cues

Adding verbal positional information to audio cues could be another means of providing effective audio cues. As suggested by Patterson (1982), speech cues can efficiently provide direct positional information that enables the listener to quickly evaluate message content. Speech cues using the words "target, target," could provide the listener with the information that the object of interest is a target and to search for that target on the visual display. Adding verbal positional information to the speech cue (i.e., "target 15 degrees") might further reduce search time by providing a more precise positional cue. In addition, spatializing the verbal cues (i.e., having "target 15 degrees" mapped to a representative location in space) could potentially further reduce the search time because spatialization provides further positional cueing, as per Perrott, Sadralodabai, Saberi, and Strybel (1991). However, no research has been conducted to

determine whether spatialized verbal positional cues would make efficient guides for visual search. In addition, no research has been conducted to determine how spatialized verbal cues would interact with different mapping schemes, especially when there is greater disparity between auditory and visual cues, such as found in the display using a ± 60 -degree auditory range. Research is needed to explore the use of verbal positional cues and the possible interaction of auditory signal mapping and verbal cues (positional or not) on aurally guided visual search.

Improved Performance Research Integration Tool (IMPRINT) Mounted Combat System (MCS) modeling data suggest that Soldiers will experience high levels of workload when performing multiple tasks, including searching visual displays. Integrating auditory display technology into the robotics collaboration OCU has the potential of reducing Soldier motor, visual, and cognitive workload by permitting him or her to more efficiently search visual displays, such as the map display. Little research has been done to explore the design of an efficient and effective audio display, even though data obtained from IMPRINT models, which indicate that audio control and display technologies would be very useful in OCU applications for tasks such as guiding visual display search (Mitchell, Samms, Glumm, Krausmann, Brelsford, & Garrett, 2004).

The purpose of this study is to examine the effects of the integration of spatial auditory display technologies in the use of visual display search tasks. Two objectives are explored within this study: (a) to determine what types of spatial audio signal mapping schemes and verbal positional cues are best for aurally guided visual display search tasks; and (b) to validate three-dimensional (3-D) audio IMPRINT models previously developed for multi-modal displays. This study will serve as the “test” phase of the “model-test-model” approach. The data used in the previous “model” phase came from journal articles and previous laboratory studies. The time and workload data obtained in this study will be compared with those previously entered into the IMPRINT MCS model using spatial audio resources related to the use of visual map displays.

In this study, it is hypothesized that the use of auditory signal mapping and verbal positional cues will reduce visual display search time. This will be useful in future experiments exploring the use of 3-D audio in demanding environments such as moving vehicles (i.e., the Army high mobility multipurpose wheeled vehicle or HMMWV), which contain relatively high levels of vibration and jolt.

2. Method

2.1 Participants

Thirty-six participants (15 males and 21 females) participated in this study. Participants had a hearing threshold level (HTL), that is, the decibel level over threshold at which the subject hears

a test stimulus, corresponding to Army physical profile H2; an average of no more than 30 dB HTL, no individual level greater than 35 HTL at 500, 1000, and 2000 Hz, and no level greater than 55 HTL at 4000 Hz, (U.S. Army, 1991). Participants had normal visual functioning (20:40 or better) and otoscopically normal ears (no blockage or infection). Participants reported no history of otologic pathology (hearing problems). Participants also had 30 hours or more experience using a computer mouse, the level at which Whisenand and Emurian (1995) considered mouse users to be “experienced.”

Participants were students attending Harford Community College (HCC), Bel Air, Maryland. They were paid \$40 for their participation in the study. The experiment was conducted in a conference room at the HCC library building.

2.2 Apparatus

Hearing tests were conducted with an Earscan² Microprocessor pure tone audiometer. Vision tests were conducted with a Snellen eye chart. Audio signals used in this experiment were synthesized into sound files by a Veridian Engineering 3-DVALS³ System II spatial audio engine, communicating with an Intel-based computer. Sound files were played on a second Intel-based computer, and participants listening to the sounds wore Sennheiser 580 headphones.

Participants viewed the visual task and listened to sound files on an Intel-based computer. The computer monitor display had an 18-inch diagonal screen. Figure 1 shows a participant with the visual and auditory task apparatus.



Figure 1. Participant with visual and auditory task apparatus.

²Earscan is a registered trademark of Micro Audiometrics Corporation.

³3-DVALS, which is a registered trademark of the Veridian Corporation, stands for 3-D virtual auditory localization system.

2.3 Stimuli

2.3.1 Visual Stimuli

The visual stimuli (visual display) consisted of U.S. Army military map symbols displayed on a color Russian topographical map, shown on the computer screen directly in front of the participant. The placement of the screen afforded the participant a 30-degree horizontal and a 26-degree vertical FOV of the visual display. On the map, Russian location names were used to discourage participant attempts to read and understand place names (English place names might potentially distract the participants from the search task). Fifty-one military map symbols were used in this study (25 U.S. Army aviation symbols, 25 Army armored cavalry symbols, and one infantry symbol to serve as the target). Fifty of these symbols appeared on the map at one time (the target symbol randomly replacing one of the other map symbols), all placed at random locations. The symbols were 0.25 inch high by 0.5 inch wide, subtending 0.6 degree of vertical visual arc and 1.2 degrees of horizontal visual arc, as per MIL-STD-1472F (1999). The location of the target was randomly changed without replacement in each trial by the software to appear at one of 31 locations between ± 15 degrees' azimuth. The center of the target was randomly located vertically between ± 13 degrees elevation. Between trials, the map symbols disappeared and a red box appeared at the center of the screen to focus the participant's gaze on the center of the screen before each trial.

2.3.2 Auditory Stimuli

The sounds heard in the experiment were pre-recorded verbal alerts played as sound files on a personal computer (PC). There were two types of audio alerts. The first was an alert without positional information, containing information regarding target occurrence with no additional azimuth location information. This alert consisted of the words, "target, target" spoken by a female voice. The second type of alert had positional information, consisting of the words, "target x degrees" spoken by the same female voice, in which " x " was the azimuth location of the visual target in degrees' azimuth.

All audio alerts had a total duration of approximately 2 seconds. The alerts were digitally pre-recorded sound files of a female talker made in a sound-treated room. These sound files were played through a Veridian Engineering 3-DVALS audio sound engine to make them correspond to the different mapping conditions. Spatialized signals used generic head-related transfer functions developed by the U.S. Air Force. All sound files were transferred to an Intel-based desktop computer for play during the experiment.

During the experiment, the participant heard the alerts through the Sennheiser 580 headphones. The root mean square (rms) amplitude of each alert, as measured under the earcup of the participant's headphones, was 78 decibels A weighted (dBA) sound pressure level (SPL). At the same time, the participant heard continuous tank noise played through the headphones. The tank noise, recorded at the commander's position in an M1A2 tank, was 65 dBA SPL measured under

the earcup of the participant's headphones. These levels were chosen because they were fairly representative SPLs for tracked vehicle noise (a noise that may be commonly found in battlefield environments), which is still within the limits of human exposure for this study. All sound and noise levels were broadband to present minimal problem to participants with hearing loss. Sound and noise were within human use hearing conservation guidelines (U.S. Army, 1991).

2.4 Procedure

Only one participant was evaluated at a time. The participant was asked to answer the volunteer agreement affidavit, undergo a vision and hearing screening (an audiogram), and receive training in all experimental tasks. After this, the experimenter read the instructions for the experimental session. Then the participant was trained to perform the visual search task.

During training and the visual search task, the participant donned the stereo headphones and sat in front of the computer monitor. Participants were seated in a chair placed 24 inches in front of a standard 18-inch diagonal computer monitor, eyes in line with the center of the screen, and were advised to keep their heads as still as possible during experimental trials. The computer screen directly in front of the participant showed the Russian topographic map with the red box in the center. No military symbols appeared on the screen. The experimenter instructed the participant to direct his or her attention to the red box at the center of the screen before each search trial. After 1.5 seconds, the red box disappeared and the computer screen showed the topographic map along with the 49 U.S. Army map symbols and one target symbol. At the same instant, the participant heard the 2-second auditory alert that described the location of the visual target. The participant was instructed to use the auditory alert as a guide to visually locate the target symbol. The participant was advised that when s/he knew where the target symbol was located, to use a computer mouse to move a cursor shown on the computer screen as quickly as possible to the location of the symbol s/he thought was the target. When the participant moved the cursor on top of the map symbol s/he had chosen, s/he was instructed to click the left-hand mouse button to indicate that s/he located the target. When the participant clicked the left-hand mouse button, the symbology disappeared and the screen with the red box reappeared to refocus the participant's gaze upon the center of the screen. Then, 1.5 seconds later, the red box disappeared and the next trial began. The cursor was automatically moved to the center of the screen at the beginning of each trial.

Before each session, the participant was trained to use the auditory mapping scheme assigned to that session and had 31 search trials as training, each trial preceeded by the screen with the red box. The target symbol appeared at all 31 azimuth angle locations (± 15 degrees azimuth, including 0 degrees) without repetition. The targets also appeared at random vertical locations within 26 degrees' elevation, so that half the targets appeared between 0 and 13 degrees' elevation, and half appeared between -13 and 0 degrees' elevation. After training, the participant performed 31 visual search trials in a data run session. For training and data run sessions, the location of the target symbols was changed for each trial by the software, and the order of

appearance was random. After the participant finished all trials in the data run session, s/he doffed the headset and completed the National Aeronautics and Space Administration Task Load Index (NASA-TLX) workload measure (Hart & Staveland, 1987). After this, the subject had a rest break of at least 5 to 20 minutes (based on participant preference) before the next session.

During the experiment, each participant participated in a total of five sessions, each lasting approximately 15 minutes. Each session involved a new auditory display-mapping scheme with the same verbal positional information used in the previous session. After the participant completed the NASA-TLX for the last experimental session, the participant had an exit audiogram, then an exit interview conducted by the experimenter to determine the mapping strategy that the participant preferred most. After this, the experiment ended and the participant was compensated. The entire experimental session, from initial orientation to final questionnaire, lasted approximately 3 hours.

2.5 Experimental Design

The treatment structure for this experiment was a mixed model factorial structure. The independent variables were as follow:

1. *Auditory mapping scheme* was a within-subject variable. The five categories of mapping scheme were
 - a. Monaural (the same auditory signal was heard in both ears);
 - b. Three spatial audio alerts, sounding at -60 , 0 , and $+60$ degrees (an audio span of $\pm 60^\circ$ degrees);
 - c. Three spatial audio alerts, sounding at -37 , 0 , and $+37$ degrees (an audio span of $\pm 37^\circ$ degrees);
 - d. Three spatial audio alerts, sounding at -15 , 0 , and $+15$ degrees (an audio span of $\pm 15^\circ$ degrees);
 - e. Visual only (no audio signal was heard).
2. *Verbal positional information* was a between-subject variable that consisted of
 - a. Audio alerts without positional information, and
 - b. Audio alerts with positional information
3. *Visual target azimuth* was a within-subject variable that was the azimuth of the visual target. Visual targets were presented at all 31 azimuth positions from -15 to $+15$ degrees, including 0 degrees.

The order of presentation of auditory mapping scheme was assigned to participants by means of a Williams Square design, which counterbalanced order of treatments to ensure that all

treatments were administered to participants without repetition in order of presentation. The order of presentation of visual target stimuli was assigned into five different random order presentations, each of which was assigned to participants in a Williams Square design. Both Williams Square designs are presented in appendix A.

Type of verbal positional information was assigned to participants at random, with the levels of this condition equally divided within the male and the female group. Each participant received only one level of this variable, and both levels of this variable were distributed equally between all participants.

The dependent variables were as follows

1. Visual target search time, which was defined as the time interval between the simultaneous triggering of the visual symbology and the auditory mapping stimulus, and the click of the mouse button indicating participant identification of the visual target; and
2. The workload rating of the target location task supplied in the NASA-TLX questionnaire.

2.6 Data Analysis

Visual target search time was analyzed in a $10 \times 5 \times 5 \times 5 \times 2 \times 2$ analysis of covariance (ANCOVA), with visual target azimuth as the covariate. Variables were Williams Square order of presentation of mapping schemes \times auditory mapping scheme \times sequential order of presentation of mapping scheme \times order of presentation of visual target symbols \times verbal positional information \times gender. Interactions between the random and fixed effect factors were used as error terms for testing hypotheses about the fixed effect factors, as prescribed by a mixed model analysis.

Workload ratings were analyzed in a $10 \times 5 \times 5 \times 5 \times 2 \times 2$ analysis of variance (ANOVA). Variables were Williams Square order of presentation of mapping schemes \times auditory mapping scheme \times sequential order of presentation of mapping scheme \times order of presentation of visual target symbols \times verbal positional information \times gender.

For target search time and workload data, effects showing a probability value p less than 0.05 were considered statistically significant. The least significant difference (LSD) *post hoc* test was performed at $p \leq 0.05$ but only if the corresponding mixed model tests on the fixed effect were statistically significant at the 0.05 level.

3. Results

3.1 Visual Target Search Time

The ANCOVA revealed significant effects for the interaction of auditory mapping scheme and verbal positional information, $F(4, 127) = 19.70, p < 0.0001$, and for the main effect of auditory mapping scheme, $F(4, 682) = 10.82, p < 0.0001$. There were also significant effects for the interaction of visual target azimuth and auditory mapping scheme, $F(4, 5365) = 5.13, p = 0.0004$, and for the main effect of visual target azimuth, $F(1, 5365) = 63.42, p < 0.0001$. There were no other significant effects. Data were for correct and incorrect target recognitions, but all subjects correctly identified almost all (99.99%) targets.

The *post hoc* analysis of the auditory mapping scheme x verbal positional information interaction is illustrated in figures 2 through 4 and is reported for absolute visual target azimuths of 0, 7.5, and 15 degrees, respectively. An absolute azimuth value of 0 degrees represents a target location at the center of the computer monitor, while absolute values of 7.5 and 15 degrees represent locations halfway across the screen and at the edge of the computer monitor, respectively, regardless of direction.

Regardless of positional information, visual target search time with spatial audio signals was significantly shorter than search times with visual only, at all azimuths. All spatial audio target search times were significantly shorter than those for monaural signals with no positional information at all azimuths. Monaural signals with positional information showed target search times no different than those with spatial audio, with or without positional information at all azimuths. Monaural signals without positional information had significantly greater target search times than monaural signals with positional information at all azimuths.

At all azimuths, the longest target search times were found for conditions in which visual data only were presented, when positional information was otherwise given in all other conditions. For conditions in which no positional information was given, visual only had significantly greater search times than any spatial audio signals, and significantly greater search times than monaural signals with positional information.

For all audio mapping and types of positional information, visual search time was shortest for targets at the center of the screen and increased as the target approached either edge of the screen. The increase with azimuth was not significant for all spatial audio signals and for monaural signals with positional information. The increase was significant for all conditions with visual only and for conditions with monaural signals containing no positional information.

The audio mapping x positional information interaction at different degrees azimuth was analyzed to determine if it precluded the interpretation of the main effect of mapping condition.

As can be seen in figures 2 through 4, at each azimuth, the monaural condition was not ordinal with respect to the other mapping conditions. Monaural audio without positional information had a greater mean than monaural audio with positional information, which is a reverse of the relationship of all the other variables. Therefore, this interaction precluded the interpretation of the mapping main effect.

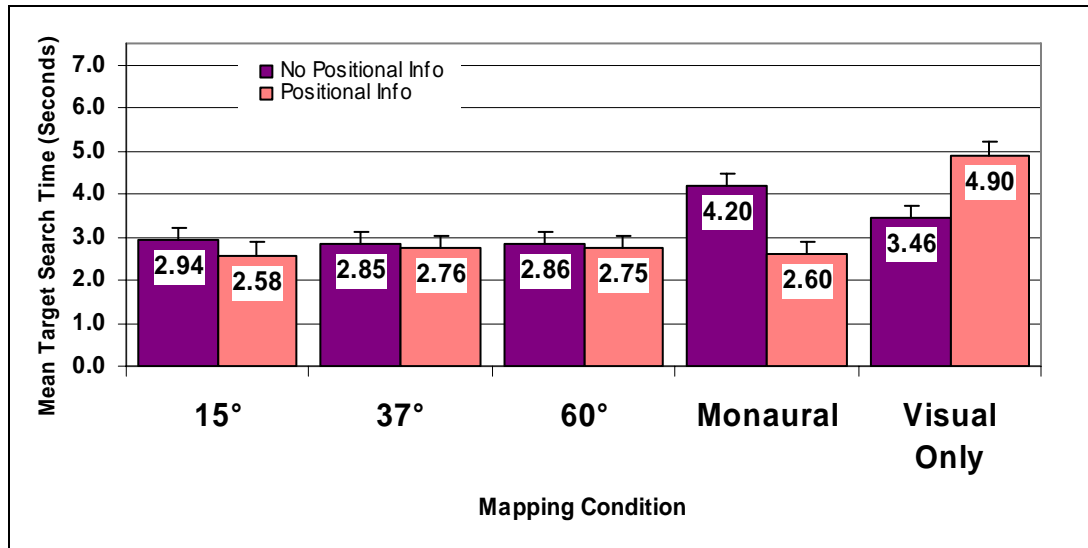


Figure 2. Mean target search times for audio mapping x positional information, visual signals at 0 degrees absolute azimuth (+ 1 standard error of the mean).

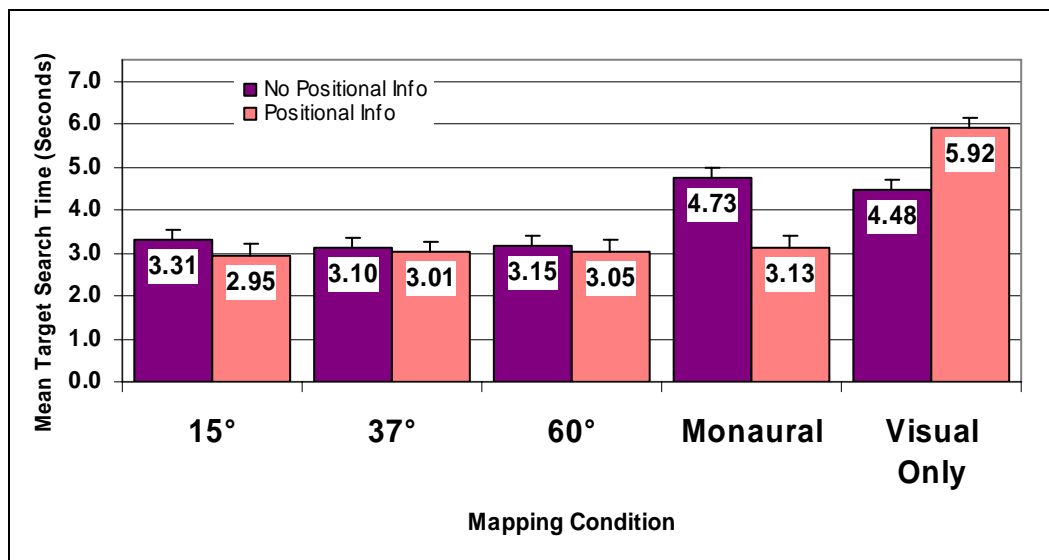


Figure 3. Mean target search times for audio mapping x positional information, visual signals at 7.74 degrees absolute azimuth (+ 1 standard error of the mean).

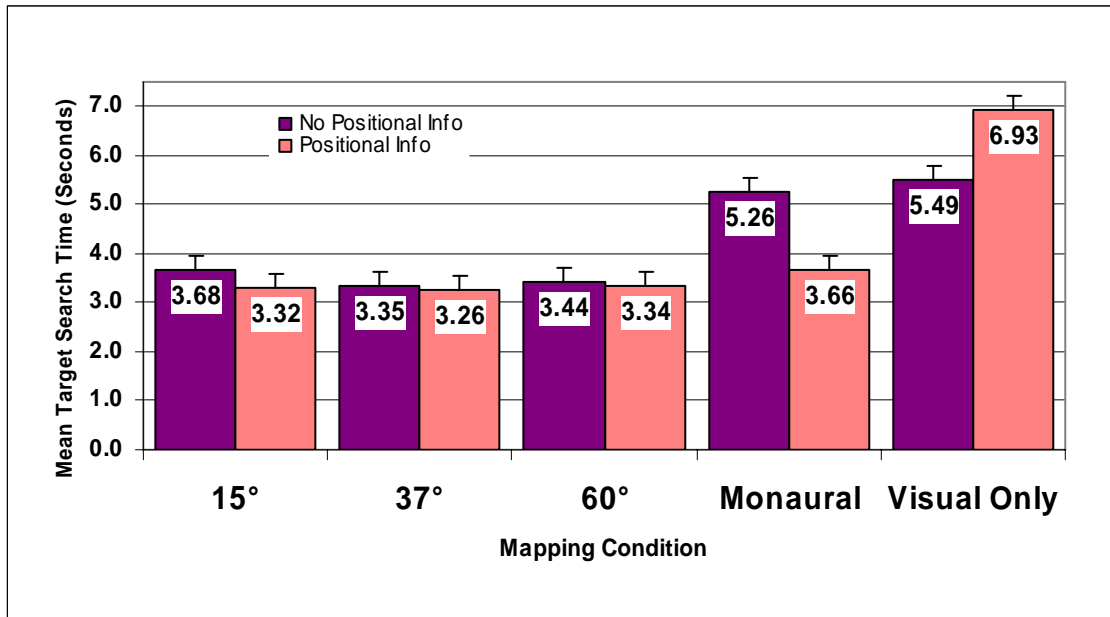


Figure 4. Mean target search times for audio mapping x positional information, visual signals at 15 degrees absolute azimuth (+ 1 standard error of the mean).

The ANOVA of workload data indicated significant effects for the interaction of auditory mapping scheme and verbal positional information, $F(4,127) = 4.08, p = 0.004$, and for the main effect of auditory mapping scheme, $F(4, 127) = 8.97, p \leq 0.0001$. *Post hoc* analysis of the auditory mapping scheme x verbal positional information interaction indicates that participant workload ratings for mapping conditions with visual only for participants assigned to the positional information condition were significantly greater than workload ratings for any other variable, regardless of positional information and mapping scheme, $p \leq 0.01$. This effect is depicted in figure 5, where it can be seen that visual only in the positional information produced workload ratings almost twice as high as those in other conditions, including visual only in positional information conditions. (The maximum possible workload score was 100.)

There were no other significant differences between means. This indicates that there were no statistically significant differences between spatial audio and monaural signals with and without positional information. There were no significant differences between all spatial and monaural signals and visual only signals in the positional information condition.

The Audio Mapping x Positional Information interaction was analyzed to determine if it precludes the interpretation of the main effect of mapping condition. As can be seen in figure 5, the monaural condition was not ordinal with respect to the other mapping conditions, meaning that the relationship between monaural with and without positional information differed in respect to the other variables. Monaural audio without positional information had a greater mean than monaural audio with positional information, which is a reverse of the relationship of all the

other variables. Therefore, this interaction precluded the interpretation of the mapping main effect.

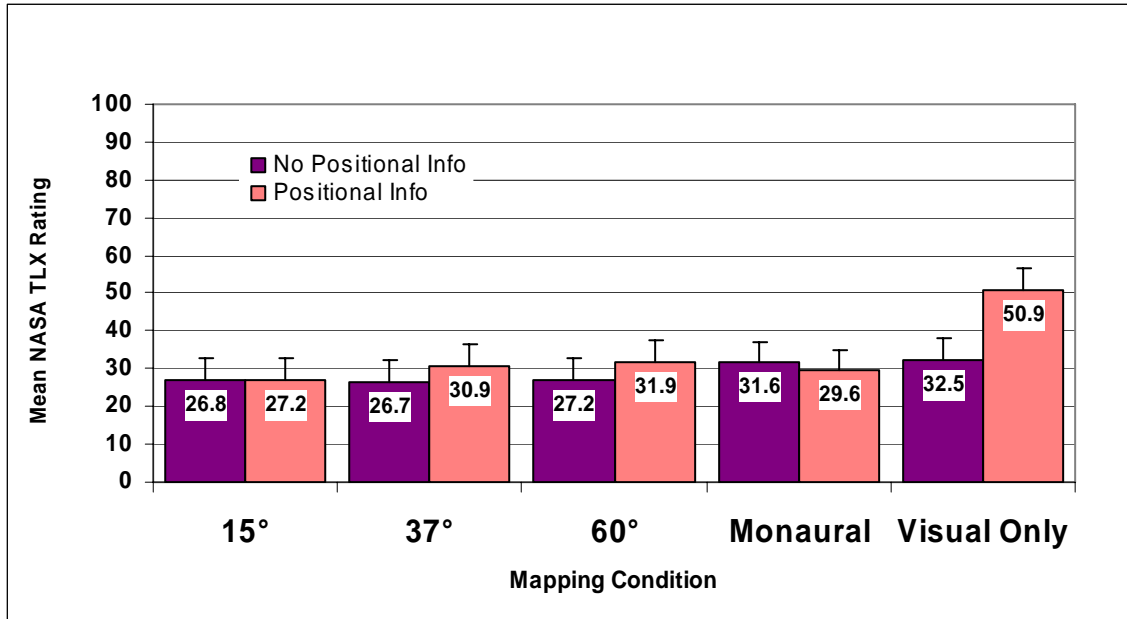


Figure 5. Mean workload ratings for Audio Mapping x Positional Information interaction (+ 1 standard error of the mean).

4. Discussion

4.1 Target Search Times

The data indicated that it was possible to determine efficient mapping and positional information schemes for displays with a narrow FOV of 31 degrees. Target search times for audio signals were found to be significantly shorter than those with visual only. These findings are in accordance with the previous findings by other researchers that auditory cues can facilitate detection of a visual target (Perrott et al., 1991; Elias, 1993, 1996; Perrott et al., 1995, 1996). These findings also agree with Begault (1993), who found that search times can be reduced when spatially correlated sounds were present. Note that spatial audio cues without positional information had search times not significantly different than those of spatial audio signals with positional information, which implies that spatial audio information alone provides efficient audio cues. However, the data also showed that target search times for monaural signals with positional information were not significantly different than those of spatial audio signals. This failed to support the hypothesis that the spatial audio schemes would produce shorter visual search times than the monaural mapping. The experimental data imply that audio cues do not

have to be spatialized to provide effective target search cues and that spoken positional information reduced search time to the same extent as monaural signals. In terms of design standards, this implies that spatial audio signals and monaural audio with positional information are best for enhancing target search. This also implies that if it is not possible to use spatialized audio signals, monaural signals with positional information can be used just as effectively. If coding audio signals with spoken positional information is not possible, the signals without positional information should be spatialized.

One hypothesis stated in this study concerned the possible effect of spatial audio cue disparity on target search time. It was hypothesized that the mapping scheme that uses the widest range of audio angles (± 60 degrees) may produce significantly shorter response times than monaural displays but longer response time than displays with a narrower range of audio angles (± 37 or ± 15 degrees) because of the greater disparity between audio and visual cues. The data indicated no such difference; target search times for the widest range of audio angles had no significantly different search times than those for the narrow range of angles. These findings are in accordance with Begault (1993, 1998). In terms of design standards, this indicates that either mapping scheme would be as effective. Future research should be conducted to determine whether more extensive ranges, such as angles of ± 90 degrees (directly to the left and to the right of the listener), could cause cue disparity that might adversely affect target search time.

Results also indicated that monaural signals without positional information were not as effective as monaural cues with positional information. In most locations on the screen, monaural signals without positional information were as ineffective as cues with visual only information, when no positional information was given for other conditions. This may have occurred because monaural signals with no positional information did not have any meaningful signal content because they did not convey any information to the participant. If the screen sometimes did not contain a target, then the monaural nonpositional signal would have been meaningful when heard (signaling the occurrence of a target event whenever it occurred), and response time to the signal might have been shorter. This implies that monaural signals are most efficient if coded with positional information and least efficient when they communicate no positional information.

At all azimuths, the longest target search times were found for conditions in which visual stimuli only were seen when positional information was given in all other conditions. Visual signals only had target search times almost twice as large as those with spatialized audio or monaural signals with positional information. Conditions in which visual only was seen and in which no positional information was otherwise given had significantly shorter target search times but were still significantly greater than those of spatial audio and monaural signals with positional information. Why did participants respond differently to visual signals only when the audio signals they heard in other conditions are heard with or without positional information? Why did participants respond differently to visual only signals with and without positional information? One could surmise that if one were exposed to audio signals with positional information, then having to find a target without that information would be more cognitively demanding because

the amount of information unavailable is greater for conditions with no positional information. However, if this were true, one would expect the results to show a significant interaction for mapping scheme x order of presentation of mapping scheme x positional information, in which participants who received the visual only condition first would have significantly shorter target search times because they would not have been exposed to the other audio conditions first. However, this was not the case; this interaction was not significant. Future research is needed to explain this finding. The results imply that visual only is the least efficient in enhancing visual search.

One might ask whether the observed performance benefits provided by auditory displays might have been the result of simply providing more information to the listener in the form of location or positional information. If information regarding the likely target location is available and can be encoded in auditory form, why could it not be encoded in some other channel (such as visual highlights or tactile cues), and would this not lead to the same performance benefit? Perhaps it could, but the main point of this study was to demonstrate how enlisting other channels such as auditory might provide value, especially when visual channels are overloaded or visual displays are over-subscribed. Future research should examine human performance in the presence of other display modalities (such as tactile displays or a combination of auditory and tactile displays) in the presence of better visual cueing in conditions in which visual channels have heavy demands placed upon them.

4.2 Workload

The finding that workload ratings for signals with visual only in the positional information condition were significantly greater than all other signals indicates that visual only is perceived as a high workload condition. The workload data are similar to those for the target search times in that signals with visual only in the positional information condition had scores almost twice as large as those for spatialized audio and monaural signals of any kind. As with the target search time data, one could guess that if the participant is exposed to audio signals with positional information and then has to search for a target in the visual only condition (without auditory cues), performing the task would be cognitively demanding because the amount of information unavailable is greater for conditions with no positional information. This is indeed reflected in the workload data. However, as with target search time data, if this were true, one would expect the results to show a significant interaction for mapping scheme x order of presentation of mapping scheme x positional information. However, this was not the case; there was no such significant interaction. Future research is needed to explain this finding. In terms of design standards, this implies that the presentation of visual only in conditions in which positional information is otherwise available is the least efficient in enhancing workload associated with visual search.

As with the target search data, the nonsignificant difference in workload ratings between spatial audio and monaural signals of any kind implies that audio cues do not have to be spatialized to

provide effective target search cues and that monaural signals with positional information reduced search time to the same extent as those without positional information. Unlike the target search data, the workload data indicated there was no significant difference between visual only in the no-positional information condition and spatial and monaural signals of any kind. This implies that the presence of visual only in no-positional information conditions has a perceived workload at the same level as those conditions with spatial and monaural mappings, with and without positional information. Future research is needed to explain this finding.

5. Conclusions

The purpose of this study was to examine the effects of the integration of auditory display technologies in visual search tasks such as those in robotic OCUs by determining what types of audio signal mapping schemes and verbal positional cues are best. The results indicate that the use of auditory signal mapping and verbal positional cues significantly reduced visual display search time and that positional cues mixed with specific audio mappings were the most efficient means of reducing visual display search time. The results of this study led to specific design recommendations regarding the use of auditory signals in environments with narrow FOV visual displays, such as OCUs for robotic systems such as MATILDA and Buster. Future research should explore why target search time is longer in visual only conditions, especially in conditions in which audio positional information is otherwise provided. Future research should also examine human performance in the presence of other display modalities (such as tactile displays) and better visual cueing in conditions in which visual channels have heavy demands placed upon them.

6. References

- Adams, S. Practical considerations for measuring Situational Awareness. *Proceedings of the Third Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment*, 157-164, 1998.
- Artez, A. J. The design of electronic map displays. *Human Factors* **1991**, 33 (1), 85-101.
- Begault, D. R. Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors* **1993**, 35 (4), 707-717.
- Begault, D. R. Virtual acoustics, aeronautics, and communications. *Journal of the Audio Engineering Society* **1998**, 46 (6), 520-530.
- Bodenhamer, A. S. *Report and comments on MANSCEN robotics projects and associated operator control units*; Unpublished report; U.S. Army Research Laboratory: Ft. Leonard Wood, MO, 2004.
- U.S. Department of Defense. *Design Criteria Standard, Human Engineering (MIL-STD 1472F)*. paragraph 5.2.6.7.4, page 38. Washington, DC, **1999**.
- Elias, B. Dynamic auditory preview for visually guided target aiming. *Proceedings of the Human Factors Society 39th Annual Meeting*, 1415-1419, 1995.
- Elias, B. The effects of spatial auditory preview on dynamic visual search performance. *Proceedings of the Human Factors Society 40th Annual Meeting*, 1227-1231, 1996.
- Fujawa, G. E.; Strybel, T. Z. The effects of cue informativeness and signal amplitude on auditory spatial facilitation of visual performance. *Proceedings of the Human Factors Society 41st Annual Meeting*, 556-560, 1997.
- Hart, S. G.; Staveland, L. E. *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*. In P.A. Hancock and N. Meshkati (Eds) *Human mental workload*. Amsterdam: Elsevier, 1987.
- Mitchell, D. K.; Samms, C. L.; Glumm, M. M.; Krausman, A. S.; Brelsford, M. L.; Garrett, L. *Improved Performance Research Integration Tool (IMPRINT) Model Analyses in Support of the Situational Understanding as an Enabler for Unit of Action Maneuver Team Soldiers Science and Technology Objective (STO) in Support of Future Combat Systems (FCS)*; ARL-TR-3405; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2004.
- Patterson, R. D. *Guidelines for auditory warning systems on civil aircraft*; (Civil Aviation Authority Paper 82017); Cheltenham, England: Civil Aviation Authority, Airworthiness Division, 1982.

- Patterson, Perrott, D. R.; Cisneros, J.; McKinley, R. L.; D'Angelo, W. R. Aurally aided detection and identification of visual targets. *Proceedings of the Proceedings of the Human Factors Society 39th Annual Meeting*, 104-108, 1995.
- Perrott, D. R.; Cisneros, J.; McKinley, R. L.; D'Angelo, W. R. Aurally aided visual search under virtual and free-field listening conditions. *Human Factors* **1996**, 38 (4), 702-715.
- Perrott, D. R.; Sadralodabai, T.; Saberi, K.; Strybel, T. Z. Aurally aided visual search in the central visual field; effects of visual load and visual enhancement of the target. *Human Factors* **1991**, 33 (4), 389-400.
- Strybel, T. Z.; Boucher, J. M.; Fujawa, G. E.; Volp, C. S. Auditory spatial cueing in visual search tasks: Effects of amplitude, contrast, and duration. *Proceedings of the Human Factors Society 39th Annual Meeting*, 109-113, 1995.
- Strybel, T. Z.; Perrott, D. R. Auditory spatial facilitation of visual search performance: parallel or serial processing? Presented at the Meeting of the Psychonomic Society, Washington, D.C, 1993.
- U.S. Army Hearing Conservation. U.S. Army Pamphlet 40-501, Washington, D.C.: Department of Defense, 1991.
- Whisenand, T. G.; Emurian, H. H. Some effects of angle of approach on icon selection. *ACM CHI95 Proceedings*, 1995.

Appendix A. Williams Square Designs Used in This Study

Table A-1. Williams square design, order of presentation of auditory mapping scheme to participants.

Participants	Session 1	Session 2	Session 3	Session 4	Session 5
1,11,21,31	37	15	60	V	M
2,12,22,32	15	V	37	M	60
3,13,23,33	V	M	15	60	37
4,14,24,34	M	60	V	37	15
5,15,25,35	60	37	M	15	V
6,16,26,36	M	V	60	15	37
7,17,27	60	M	37	V	15
8,18,28	37	60	15	M	V
9,19,29	15	37	V	60	M
10,20,30	V	15	M	37	60

M = Monaural (the same auditory signal is heard in both ears)

60 = Spatial audio alerts given from 3 positions within a range of ± 60 degrees

37 = Spatial audio alerts given from 3 positions within a range of ± 37 degrees

15 = Spatial audio alerts given from 3 positions within a range of ± 15 degrees

V = Visual Stimuli Only (no audio cues)

Table A-2. Williams square design, order of presentation of visual target stimuli to participants.

Participants	Session 1	Session 2	Session 3	Session 4	Session 5
1,11,21,31	C	D	B	E	A
2,12,22,32	D	E	C	A	B
3,13,23,33	E	A	D	B	C
4,14,24,34	A	B	E	C	D
5,15,25,35	B	C	A	D	E
6,16,26,36	A	E	B	D	C
7,17,27	B	A	C	E	D
8,18,28	C	B	D	A	E
9,19,29	D	C	E	B	A
10,20,30	E	D	A	C	B

A through E are five different randomization schemes

NO. OF
COPIES ORGANIZATION

1
(PDF
ONLY) DEFENSE TECHNICAL
INFORMATION CTR
DTIC OCA
8725 JOHN J KINGMAN RD
STE 0944
FORT BELVOIR VA 22060-6218

1 US ARMY RSRCH DEV & ENGRG CMD
SYSTEMS OF SYSTEMS
INTEGRATION
AMSRD SS T
6000 6TH ST STE 100
FORT BELVOIR VA 22060-5608

1 INST FOR ADVNCD TCHNLGY
THE UNIV OF TEXAS AT AUSTIN
3925 W BRAKER LN STE 400
AUSTIN TX 78759-5316

1 DIRECTOR
US ARMY RESEARCH LAB
IMNE ALC IMS
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 DIRECTOR
US ARMY RESEARCH LAB
AMSRD ARL CI OK TL
2800 POWDER MILL RD
ADELPHI MD 20783-1197

2 DIRECTOR
US ARMY RESEARCH LAB
AMSRD ARL CS OK T
2800 POWDER MILL RD
ADELPHI MD 20783-1197

3 US ARMY NATICK SOLDIER CTR
ERGONOMICS TEAM/SSTD
ATTN J B SAMPSON
100 KANSAS ST
NATICK MA 01760-5020

1 RDECOM-NATICK SOLDIER CTR
MANAGEMENT TEAM/PEO SOLDIER/PM
FUTURE FORCE WARRIOR TECH PRO OFC
ATTN C FITZGERALD
100 KANSAS ST
NATICK MA 01760-5020

NO. OF
COPIES ORGANIZATION

1 RDECOM-NATICK SOLDIER CTR
HUMAN PERF/MANPRINT/SSI LEAD
FUTURE FORCE WARRIOR TECH PRO OFC
ATTN C L BLACKWELL
100 KANSAS ST
NATICK MA 01760-5020

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR M DR M STRUB
6359 WALKER LANE SUITE 100
ALEXANDRIA VA 22310

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MA J MARTIN
MYER CENTER RM 2D311
FT MONMOUTH NJ 07703-5630

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MC A DAVISON
320 MANSCEN LOOP STE 166
FT LEONARD WOOD MO 65473-8929

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MD T COOK
BLDG 5400 RM C242
REDSTONE ARSENAL AL 35898-7290

1 COMMANDANT USAADASCH
ATTN ATSA CD
ATTN AMSRD ARL HR ME MS A MARES
5800 CARTER RD
FT BLISS TX 79916-3802

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MI J MINNINGER
BLDG 5400 RM C242
REDSTONE ARSENAL AL 35898-7290

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MM DR V RICE
BLDG 4011 RM 217
1750 GREELEY RD
FT SAM HOUSTON TX 78234-5094

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MG R SPINE
BUILDING 333
PICATINNY ARSENAL NJ 07806-5000

1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MH C BURNS
BLDG 1002 ROOM 117
1ST CAVALRY REGIMENT RD
FT KNOX KY 40121

NO. OF
COPIES ORGANIZATION

- 1 ARMY RSCH LABORATORY - HRED
AVNC FIELD ELEMENT
ATTN AMSRD ARL HR MJ D DURBIN
BLDG 4506 (DCD) RM 107
FT RUCKER AL 36362-5000
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MK MR J REINHART
10125 KINGMAN RD
FT BELVOIR VA 22060-5828
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MV HQ USAOTC
S MIDDLEBROOKS
91012 STATION AVE ROOM 111
FT HOOD TX 76544-5073
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MY M BARNES
2520 HEALY AVE STE 1172 BLDG 51005
FT HUACHUCA AZ 85613-7069
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MP D UNGVARSKY
BATTLE CMD BATTLE LAB
415 SHERMAN AVE UNIT 3
FT LEAVENWORTH KS 66027-2326
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR M DR B KNAPP
ARMY G1 MANPRINT DAPE MR
300 ARMY PENTAGON ROOM 2C489
WASHINGTON DC 20310-0300
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MJK MS D BARNETTE
JFCOM JOINT EXPERIMENTATION J9
JOINT FUTURES LAB
115 LAKEVIEW PARKWAY SUITE B
SUFFOLK VA 23435
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MQ M R FLETCHER
US ARMY SBCCOM NATICK SOLDIER CTR
AMSRD NSC SS E BLDG 3 RM 341
NATICK MA 01760-5020
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MT DR J CHEN
12350 RESEARCH PARKWAY
ORLANDO FL 32826-3276

NO. OF
COPIES ORGANIZATION

- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MS MR C MANASCO
SIGNAL TOWERS RM 303A
FORT GORDON GA 30905-5233
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MU M SINGAPORE
6501 E 11 MILE RD MAIL STOP 284
BLDG 200A 2ND FL RM 2104
WARREN MI 48397-5000
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MF MR C HERNANDEZ
BLDG 3040 RM 220
FORT SILL OK 73503-5600
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MW E REDDEN
BLDG 4 ROOM 332
FT BENNING GA 31905-5400
- 1 ARMY RSCH LABORATORY - HRED
ATTN AMSRD ARL HR MN R SPENCER
DCSFDI HF
HQ USASOC BLDG E2929
FORT BRAGG NC 28310-5000
- 1 US ARMY CECOM-NVESD
ATTN AMSEL RD NV LWD D MOODY
10221 BURBECK RD
FORT BELVOIR VA 22060-5806

ABERDEEN PROVING GROUND
- 1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL CI OK TECH LIB
BLDG 4600
- 1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL CI OK S FOPPIANO
BLDG 459
- 2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL HR MR F PARAGALLO
AMSRD ARL HR S L PIERCE
BLDG 459
- 10 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL HR SC E HAAS
BLDG 459